AFRL-RH-WP-TR-2010-0063

Comparing Manual and Cooperative Control Mission Management Methods for Wide Area Search Munitions

Lamar Warfield
Thomas R. Carretta
Michael J. Patzek
Warfighter Interface Division
Supervisory Control Interfaces Branch
Wright-Patterson AFB OH 45433-7511

Justin R. Estepp
Warfighter Interface Division
Collaborative Interfaces Branch
Wright-Patterson AFB OH 45433-7022

John K. O'Neal
Advanced Guidance Division
Weapons Dynamics and Controls Sciences Branch
Eglin AFB FL 32542-5427

November 2009

Interim Report for July 2007 to October 2008

Approved for public release; distribution is unlimited.

Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Warfighter Interface Division
Supervisory Control Interfaces Branch
Wright-Patterson AFB, OH 45433

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2010-0063 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//
Allen J. Rowe, Work Unit Manager
Supervisory Control Interfaces Branch

//signed//
Michael A. Stropki, Chief
Warfighter Interface Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE	*Form Approved* OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 18-11-2009	**2. REPORT TYPE** Interim	**3. DATES COVERED** *(From - To)* July 2007 – October 2008

4. TITLE AND SUBTITLE Comparing Manual and Cooperative Control Mission Management Methods for Wide Area Search Munitions	**5a. CONTRACT NUMBER** In-house
	5b. GRANT NUMBER NA
	5c. PROGRAM ELEMENT NUMBER 62202F
6. **AUTHOR(S** Lamar Warfield, Thomas R. Carretta, Michael J. Patzek, Justin R. Estepp, & John K. O'Neal	**5d. PROJECT NUMBER** 7184
	5e. TASK NUMBER 09
	5f. WORK UNIT NUMBER 71840919
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**PERFORMING ORGANIZATION REPORT NUMBER** NA
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory 711th Human Performance Wing Human Effectiveness Directorate Warfighter Interface Division Supervisory Control Interfaces Branch Wright-Patterson AFB OH 45433-7511	**SPONSOR/MONITOR'S ACRONYM(S)** 711 HPW/RHCI **11. SPONSORING/MONITORING AGENCY REPORT NUMBER** AFRL-RH-WP-TR-2010-0063

12. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

88ABW/PA Cleared 05/11/2010; 88ABW-2010-2575.

14. ABSTRACT

Wide Area Search Munitions (WASMs) combine the attributes of unmanned aerial vehicles with those of traditional munitions. The WASM concept envisions artificially intelligent munitions that communicate and coordinate with one another and with human operators to effectively to perform their tasks. This study examined target acquisition for unaided operators with that of an automated cooperative controller for a complex task involving the prosecution of ground-based targets. Participants completed nine trials for each control mode (manual and cooperative) by number of WASMs (4, 8, or 16) combination. Target hit rate was not affected by control mode or number of WASMs; however, target acquisition efficiency degraded under manual control and as the number of WASMs increased. Workload was greater for the manual mode and increased as the number of WASMs increased. Self-ratings of the ability to perform a simultaneous attack were lower for the manual mode and decreased as the number of WASMs increased.

15. SUBJECT TERMS

Target acquisition, cooperative control, wide area search munitions, WASMs

16. SECURITY CLASSIFICATION OF: Unclassified			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Allen Rowe
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UNL	46	**19b. TELEPONE NUMBER (*Include area code)***

This page intentionally left blank.

CONTENTS

FIGURES

TABLES

GLOSSARY

3 DoF	3 degree of freedom
711 HPW/RHCI	711th Human Performance Wing, Supervisory Control Interfaces Branch
AFRL/RWGN	Air Force Research Laboratory, Munitions Directorate
ANN	artificial neural network
CONEMP	concepts of employment
CSIL	Crew Systems Integration Laboratory
ECG	electrocardiogram
EEG	electroencephalogram
FFT	Fast Fourier Transform
GUI	graphical user interface
HEOG	horizontal electrooculogram
LOCASS	Low Cost Autonomous Attack System
NASA TLX	National Aeronautics and Space Administration Task Load Index
NuWAM	workload assessment monitor
SA	situaional awareness
TAI	tactical areas of interest
TSD	tactical situation display
UAV	unmanned aerial vehicle
USAF	United States Air Force
WASM	wide area search munitions

This page intentionally left blank.

PREFACE

This report describes activities performed in support of the Air Force Research Laboratory Warfighter Interface Division, Supervisory Control Interfaces Branch (711HPW/RHCI) Interfaces for Small Unmanned Systems, Work Unit 71840917. The authors thank Sarah E. Spriggs and Airam Gonzalez-Garcia (711 HPW/RHCI) for computer support**,** Dr. Guy French (711 HPW/RHCI) for assistance in development of test materials and procedures, and William D. Miller and Jason W. Monnin (711 HPW/RHCP) for their assistance with physiological data collection. Finally, we thank Doug Zimmer and Mike Martin of Lockheed for software development support.

This page was intentionally left blank.

1.0 INTRODUCTION

Future unmanned systems are expected to be more autonomous than those that are currently operational. In these future systems, a single operator may be expected to monitor and exert executive control over several unmanned systems (Barbato, 2000; Clough, 2002; Prieditis, Dalal, Arcilla, Groel, Van Der Bock, & Kong, 2004). The United States Air Force (USAF) is considering advanced automation system concepts that could deploy multiple semi-autonomous unmanned weapons systems into the battle zone. One such system, the Wide Area Search Munitions (WASMs), is a hybrid that combines the attributes of an unmanned aerial vehicle (UAV) such as loiter and surveillance with those of traditional fly-over-shoot-down and hit-to-kill munitions. The WASM concept envisions semi-autonomous, intelligent munitions that communicate and coordinate with one another and human operators to search Tactical Areas of Interest (TAI) and engage ground targets encountered within the TAIs. Preliminary Concepts of Employment (CONEMP) developed by operators during testing at the 19th Special Operations Squadron, Hurlburt Field FL called for WASMs to be deployed individually or in groups from larger aircraft. Cooperative control concepts have been proposed to enhance coordination among the WASMs leading to optimal resource allocation (Goraydin, 2003; Scerri, Liao, Lai, Sycara, Xu, & Lewis, 2004; Schumacher, Chandler, & Rasmussen, 2002; Schumacher, Chandler, Rasmussen, & Walker, 2003). Automation in the form of a centralized cooperative control algorithm, as well as on-board automatic target acquisition technology, has been included in these system concepts to help in path planning for search and synchronized routing and attack for coordinated rendezvous. Research into strategies for controlling the WASMs presents a challenge that is being approached by simulating the WASMs using a 3 Degree of Freedom (3 DoF) simulation and evaluating the performance when integrated with human-in-the-loop simulations and CONEMP scenarios. The Low Cost Autonomous Attack System (LOCAAS) was the first generation of such search munitions and served as the basis for the WASM testbed used to conduct human-in-the-loop simulations. Algorithms have been developed to evaluate the ability to simultaneously deploy 200 WASMs to search and destroy ground based targets in a coordinated support role with manned aircraft (Scerri, Liao, Lai, Sycara, Xu,

& Lewis, 2004). Specifically, questions regarding the operator's Situational Awareness (SA) and workload and the overall impact on target acquisition needs to be researched and developed. In addition, the level of autonomy that will reside in the munition as opposed to the operator needs to be addressed as well. This level of "tunable autonomy" will be dependent upon the operator's ability to multitask and was investigated as part of testing conducted by the Warfighter Interface Division.

The objective of the current study was to examine target acquisition performance for unaided human operators with that of an automated cooperative controller in accomplishing a complex task involving the prosecution of ground based targets with WASMs. The purpose of the study was to provide empirical data on a human's ability to simultaneously manage multiple WASMs while performing a target search, identification, and weapon assignment task. This information will provide valuable insights into concepts of employment and technology requirements for future munitions and semi-autonomous systems and aid in addressing the "tunable autonomy" problem. (e.g., how much automation is acceptable, information requirements, need for decision aiding software, manpower and personnel qualification requirements).

2.0 METHOD

2.1. Participants

Twelve full-time civilian and military employees stationed at Wright-Patterson AFB OH participated in this study. This sample consisted of 12 men who ranged in age from 20 to 45 years with a mean of 30.3 years. All participants reported being in good to excellent health and having vision correctable to 20/20, normal color vision, and normal peripheral vision. Most participants indicated that they had previous simulator (67%) and video game (92%) experience. Participation was voluntary and no compensation was offered in exchange for participation in this study.

2.2. Weapons Assignment Task

A 2 by 3 within-subjects experimental design was used. The two independent variables were level of control for planning the attack (manual or cooperative control mode) and number of WASMs launched (4, 8, or 16).

Participants were seated in front of a laptop computer that depicted a tactical situation display (TSD). The TSD used FalconView maps that contained icons for the targets and the WASMs. Participants used the TSD to view the projected flight paths and the locations of the WASMs, and targets. Participants were told the suspected number of targets in the area to be prosecuted and the priority, high or low, of each target type. Participants were instructed to attack only those items whose LADAR images match those of the assigned high and low priority targets.

At the beginning of each trial, participants launched the appropriate number of WASMs (4, 8, or 16) for that trial. The WASMs were launched in groups of 4. As the trial progressed, participants were required to review LADAR imagery from the WASMs and make target identification decisions (high priority target, low priority target, or non-target). Once the targets were identified, participants assigned the WASMs to the targets.

In the manual control mode, participants assigned a WASM to a target by first clicking the mouse on the WASM, then dragging a line from the WASM to the target of interest. After the target and the WASM were shown to be connected by a line, a GUI would appear and prompt the participant as to whether they wanted to "attack" the designated target. To authorize the attack, participants clicked "OK" using the mouse. In the manual control mode, the prosecution of the target would begin immediately following authorization for that target.

In the cooperative control mode, participants clicked on each target(s) to be attacked. As targets were chosen, they were added to a list in a GUI, which would also keep track of the number of targets designated. After all the desired targets were designated for attack, the participant would authorize the cooperative controller to develop a proposed flight path solution. Once the cooperative controller indicated a proposed flight path solution, the participant had approximately 10 seconds to review the result and either accept or reject the solution.

2.3. Measures

Three types of data were collected – task performance measures, physiological measures, and questionnaires.

2.3.1. Task Performance Measures

Several objective measures of target acquisition performance were collected. These were number of high priority targets attacked, number of low priority targets attacked, mean time on target, mean time on target error, standard deviation of time on target, time to plan, and time to complete. Brief definitions of each of these measures are provided in Table 1.

Table 1. Objective Measures of Task Performance

Measure	Definition
Number of High Priority Targets Attacked	Mean number of high priority targets attacked
Number of Low Priority Targets Attacked	Mean number of low priority targets attacked
Mean Time on Target	The average time on target for the WASMs This is the average time from launch of the WASMs till detonation of the targets.
Mean Time on Target Error	The average error between the time on target and requested time on target. That is, how close the attacks were to the requested time. This score could be computed only for the cooperative control condition.
Standard Deviation of Time on Target	This is the standard deviation of the actual time on target compared with mean time on target (i.e., how close the attacks were to each other).
Time to Plan	Time from when the first target was selected to attack authorization or cancellation.
Time to Complete	Time from authorization to when the last target is attacked.

2.3.2. Physiological Measures

While the task was being performed, physiological data was collected from the participant to obtain objective measures of cognitive workload (as previously demonstrated by Wilson & Russell, 2003a, 2003b). Electrical brain activity (electroencephalogram, or EEG) was recorded from five electrode sites (Fz, F7, Pz, T5 and O2) according to the International 10-20 system for electrode placement (Jasper, 1958) using an appropriately-sized electrode cap (Electro-Cap International, Inc., Eaton, OH) with 9 [mm] embedded tin disk electrodes (see Figure 1). A monopolar reference montage was used; each scalp site was referenced to a single 9 [mm] tin cup electrode (Electro-Cap International, Inc., Eaton, OH) placed on the left mastoid process. Vertical eye activity (vertical electrooculogram, or VEOG) and horizontal eye activity (horizontal electrooculogram, or HEOG) were monitored from bipolar pairs of Ag/AgCl electrodes (Pediatric Huggables; ConMed Corporation, Utica, NY) placed above and below the left eye and outside the outer canthus of each eye, respectively. Electrical heart activity (electrocardiogram, or ECG) was monitored using a bipolar electrode pair (Pediatric Huggables) placed on the participant's sternum and left clavicle. A Cleveland Medical Devices BioRadio 110 telemetry device (with amplifier ground from a single 9 mm tin cup electrode on the right mastoid process) was used to amplify, digitize (with a sampling rate of 200 Hz), filter (band-pass filtered between 0.5 and 52.4 Hz) and wirelessly transmit the data to custom-developed software suite for real-time physiological data collection and processing, NuWAM (Workload Assessment Monitor; Krizo, Wilson, & Russell., 2005). The physiological data was saved for offline, post-hoc analysis. Before data collection began, the electrode impedance at each EEG site was determined to be below 5 kohms.

2.3.3. Study Questionnaires

The questionnaires were a Demographic Data/Background Questionnaire, Confidence Ratings, the National Aeronautics and Space Administration Task Load Index (NASA-TLX: Hart & Staveland, 1988), and a Post-Test Questionnaire. The questionnaires used in this study are described below and provided in Appendix A.

***Figure 1*. The WASM experimental station in the CSIL facility.**

Demographic data/background questionnaire. This questionnaire was used to collect information in order to characterize the participant's prior experience and demographic characteristics and assist in interpretation of participant's performance on the target acquisition/weapons assignment task. Items elicited information about participant's sex, age, general health, vision (i.e., correctable to 20/20 acuity, normal color and peripheral vision), wellbeing, experience with simulator-type environments, video game experience.

Confidence ratings. At the completion of each target acquisition/weapon assignment scenario, participants were instructed to indicate the level of confidence in their target acquisition decisions. Confidence ratings were made on a five-point Likert rating scale (1 - not at all confident, 2 - slightly confident, 3 - moderately confident, 4 - fairly confident, 5 – very confident).

NASA-Task Load Index (NASA-TLX). The NASA-TLX (Hart & Staveland, 1988) is a subjective workload assessment measure that allows users to evaluate their interactions with human-machine systems. A computerized version of the tool was used

to assess operator workload during this study. The NASA TLX uses a multidimensional weighting procedure to derive an overall workload score based on weighted averages of ratings on 6 subscales: Mental, Physical, Temporal, Effort, Performance, and Frustration. Definitions for each of the subscales. See Figure A-1

After performing each test trial, raters marked each subscale at the point that reflected their experience. Each subscale has 20 points and the endpoints are labeled Low and High (the endpoints for Performance are Good and Poor). After completion of the subscale ratings, raters were required to make a series of pairwise comparisons among the subscales to determine their relative contributions to overall workload. The subscale ratings are weighted according to their subjective importance to raters performing a specific task. Ratings of factors deemed most important in contributing to the workload of a task are given more weight in computing the overall workload score. Subscale scores and Total workload scores may range from 0 to 100.

Post-test questionnaire. This questionnaire elicited information regarding participants' assessment of the operator interface. Participants rated the operator interface for ease of use to identify the targets and classify their priority level (high or low). Participants also provided a self-assessment of their ability to perform a near simultaneous attack under the manual and cooperative control conditions for the 4 and 16 WASM scenarios. These questions used a five-point scale (1 - poor, 2 - fair, 3 - good, 4 - very good, 5 – excellent). Participants also were given the opportunity to provide comments regarding the operator interface and other factors that affected their ability to identify, classify, and attack targets.

2.4. Equipment

The study was conducted in the Crew Systems Integration Laboratory (CSIL) in the 711th Human Performance Wing, Supervisory Control Interfaces Branch (711 HPW/RHCI) using the WASM testbed, provided by the Munitions Directorate (AFRL/RWGN). As shown in Figure 1, participants were seated in a crew member's chair attached to rails. The chair was located in the aft end of a generic cargo aircraft simulator. Participants were seated directly in front of a 13.3 inch CF-73 Panasonic laptop that presented the simulated wide area search munitions attacking targets on a

FalconView map. Still images of potential targets were displayed on a poster next to the laptop computer to aid the participants during target acquisition. Participants used a mouse with a scroll wheel to designate targets and make weapon assignments. A laptop computer was placed nearby where participants entered questionnaire responses.

Using the BioRadio 110 telemetry system, the physiological data was transmitted wirelessly from a transmitter inside the aircraft simulator to a receiver at a workstation outside of the simulator. The receiver was connected to a laptop PC running the NuWAM software.

2.5. Procedures

The experimental session began with a pre-briefing, participant informed consent, and completion of a short biographical questionnaire. The pre-briefing provided information regarding the purpose of the study, equipment, controls, and displays to be used, procedures, and the mission scenario. As part of the pre-briefing, participants were informed that they could withdraw from the study at any time and discontinue further participation without prejudice. Participants remained in a seated position during the pre-briefing, practice, and data collection.

Following the pre-briefing, training was conducted to achieve familiarity with test equipment, procedures, and tasks. Participants completed three practice trials for each level of control (manual vs. cooperative control) by number of WASMs (4, 8, or 16) combination using a representative target set. Prior to starting the test trials, participants were fitted with the electrodes for physiological measurement.

There were nine test trials for each level of control (manual or cooperative control) by number of WASMs (4, 8, or 16) combination. Level of control was randomized between participants. Within each level of control, participants first completed all 4 WASM trials, followed by the 8 WASM trials, then the 16 WASM trials. Immediately following each test trial, participants rated the level of confidence in their target acquisition decisions and subjective workload. Each trial was approximately 5-10 minutes in length (with some variation due to number of WASMs and whether the level of control was manual or cooperative), and between-trial breaks of 5-10 minutes were used to reset the simulation and data collection software for the next trial. A lunch break

of approximately one hour was taken after the first 9 trials (half of the total number of trials) were completed. After completion of the final test session, participants completed the post-test questionnaire regarding their experience. Finally, the electrodes were removed from the participant. In total time (including the lunch break), data collection for a single participant lasted between 6 and 7 hours over the course of a single duty day.

2.6. Analyses

2.6.1. Task Performance Measures

The purpose of the study was to compare the objective and subjective data on a target acquisition task for manual versus cooperative control over three levels of mission complexity (4, 8, or 16 WASMs). Related samples t-tests and repeated measures analyses of variance were performed since participants were exposed to all levels of control by number of WASMs combinations. Partial eta squared and observed power were reported in conjunction with the analyses of variance. Partial eta squared is a measure of effect size. It is the proportion of the effect plus error variance that is attributable to the effect; thus, the larger the value the more variance that is explained by the effect (e.g., level of control, number of WASMs). The power of a statistical test is the probability that the test will reject a false null hypothesis (will not make a Type II error). As power increases, the probability of a Type II error decreases. Observed power is conducted after the study has been completed and used the obtained sample size and effect size to determine what the power was in the study, assuming the effect size in the study is equal to that in the population. As with partial eta squared, the larger the value the better.

Objective measures of performance included number of hits, number of false alarms, and several target acquisition efficiency scores (see Table 1). Subjective measures were overall workload, confidence in target acquisition decisions, and their self-assessment of the ability to accomplish near simultaneous attack.

It was assumed that task difficulty would increase going from cooperative control to manual control and as the number of WASMs increased from 4 to 8 to 16. As a result, all analyses were performed using a directional hypothesis. A .05 Type I error rate was used for all analyses.

2.6.2. Physiological Measures

In addition to the subjective workload data, an objective measure of workload was derived from the physiological data using methods similar to those presented in Wilson & Russell (2003a, 2003b). Using the physiological data (sampled at 200 Hz, band-pass filtered between 0.5 and 52.4 Hz), 36 features were created that were used subsequently in a pattern recognition paradigm. These features included inter-beat interval from the ECG data (using R-R interval), as well as band power from the five traditional EEG bands: Delta, 1-3 Hz; Theta, 4-7 Hz; Alpha, 8-12 Hz; Beta, 13-31 Hz; and Gamma, 31-43 Hz. The band power features were calculated over a 1 s window via a Fast Fourier Transform (FFT) with a 200-point Hamming window (given that the sampling rate was 200 Hz). All 36 features were smoothed using a 10 s window with a 9 s overlap, resulting in a final time resolution of 1 s.

These 1 s feature vectors were used to train a fully-connected, feed-forward artificial neural network (ANN) classifier, with back propagation training, to discriminate between two classes in a training set. The specific ANN used the 36 input features, with a hidden layer of 36 nodes, and two output nodes (for the two-class training set). For each training set, the number of exemplars from each of the two classes was balanced. From this balanced training set, 75% of the available data was used for ANN training, while the remaining 25% was used as a validation set to prevent over-learning of the training set.

To objectively determine the operator’s workload for each type of control (manual vs. cooperative) over the three levels of mission complexity (4, 8, or 16 WASMs), the ANN was trained to the two-class problem of discriminating between the lowest workload condition and the highest workload condition. Analysis of the subjective NASA-TLX workload data confirmed that these two conditions, across the group of 12 participants, were 4 WASM/Cooperative and 16 WASM/Manual, respectively. Only data from the target identification and weapon assignment portions of the task were used in the training sets, since these were the only portions of the task that required the operator to be cognitively engaged with the simulation. Similarly, only data from these two subtasks were used in test sets where the trained ANN was used to discriminate between high and low workload. By training the ANN to discriminate between the lowest workload and highest workload conditions, using test data sets from all six conditions

(manuals or cooperative control over 4, 8, or 16 WASMs) provides an objective measurement of the level of cognitive demand required by each level of task.

To optimize ANN classification performance, a top-down feature selection method was implemented to determine the feature set that yielded the best classification results for each trained network. Beginning with the largest feature set (36 features), ANN training was iterated by removing the least salient feature from the previous feature set, and re-training on the new (n-1) feature set. The ranked saliency of each feature was determined using the Ruck method (Ruck, Rogers, & Kabrisky, 1990). Each trained ANN was tested on a dataset independent from the training set. For the two training conditions (4 WASM/Cooperative and 16 WASM/Manual), the validation set (25% of the data withheld from the training set to prevent over-learning) was used as the test set. For the remaining four conditions, the test sets consisted of all the data collected from the target identification and weapon assignment subtasks. Results of the ANN's performance on each test set were reported as the percentage of the test data the ANN classified as being from the low workload condition (4 WASM/Cooperative). Estimated class for each sample vector was assigned according to which output node of the network was calculated to have the higher weight.

3.0 RESULTS

3.1. Target Acquisition Performance

3.1.1. Number of Hits and False Alarms

It was expected that performance under the cooperative control mode would equal or exceed that under the manual control mode, so one-directional hypotheses were tested. Comparisons between the cooperative control and manual control modes indicated that within each number of WASMs condition, there was no significant decrement in the number of high priority targets attacked. However, for the number of low priority targets attacked, more targets were attacked for the cooperative control mode in the 16 WASMs condition (3.69 vs. 3.41; $t = 2.41$, $p \leq .05$). See Table 2 for a summary of the related samples t-tests. Although we intended to examine number of false alarms, we were unable to because the rate was extremely low with only 2 false alarms across all

participants.

Table 2. Number of Hits: Cooperative Control versus Manual Mode

Score	N WASMs	Cooperative Control Mean	Cooperative Control SD	Manual Mean	Manual SD	df	t
N High Priority Hits	4	3.33	0.00	3.27	0.12	11	1.48
	8	6.66	0.00	6.55	0.38	11	1.00
	16	12.30	0.09	12.52	0.33	11	-2.00
N Low Priority Hits	4	0.66	0.00	0.69	0.17	11	-0.56
	8	1.33	0.00	1.33	0.14	11	0.00
	16	3.69	0.09	3.41	0.35	11	2.41*

N = 12; *p≤ .05

3.1.2. Time on Target, Time to Plan, and Time to Complete Measures

Means and standard deviations for the time on target, time to plan, and time to complete measures are presented in Table 3. It should be noted that mean time on target error (i.e., average error between the actual time on target and requested time on target) cannot be computed for the manual mode because a requested time on target cannot be specified in manual mode.

Results for the repeated measures analyses of variance are summarized in Table 4. For the *Mean Time on Target* score, no significant effects were observed for level of control, number of WASMs, or their interaction. *Mean Time on Target Error* (i.e., how close the attacks were to the requested time) generally increased as the number of WASMs/targets increased (4 WASMs = 2.04, 8 WASMs = 1.30, 16 WASMs = 8.58). The low value for the 8 WASM condition may have occurred due to the closer placement of targets in this condition relative to the 4 WASM/targets condition.

Table 3. Means and Standard Deviations: Time on Target, Time to Plan, and Time to Complete Scores`

		Cooperative Control		Manual	
Score	N WASMs	Mean	SD	Mean	SD
Mean Time on Target	4	494.00	83.88	573.84	327.90
	8	488.57	55.83	446.71	67.35
	16	540.15	75.55	552.56	288.37
Mean Time on Target Error	4	2.04	1.22	-------	-------
	8	1.30	0.53	-------	-------
	16	8.58	4.44	-------	-------
SD Time on Target Error	4	2.24	2.11	10.17	4.21
	8	1.45	1.44	17.58	7.16
	16	9.09	6.16	27.43	11.89
Time to Plan	4	22.47	4.00	39.40	15.66
	8	36.01	7.63	61.26	26.83
	16	70.16	13.71	105.24	51.05
Time to Complete	4	117.22	11.89	63.06	10.45
	8	124.63	7.49	65.64	5.43
	16	148.09	26.76	74.96	10.90

N = 12

Table 4. Repeated Measures Analysis of Variance Multivariate Test Results: Time on Target, Time to Plan, and Time to Complete Scores

Score	Effect	df1	df2	F	Partial Eta^2	Observed Power
Mean Time on Target	Level of Control	1	11	0.13	0.012	0.063
	N WASMs	2	10	2.80	0.380	0.429
	Level of Control by N WASMs	2	10	1.18	0.191	0.203
Mean Time on Target Error	N WASMs	2	10	6.96*	0.582	0.822
SD Time on Target Error	Level of Control	1	11	40.69**	0.787	1.000
	N WASMs	2	10	49.63**	0.908	1.000
	Level of Control by N WASMs	2	10	11.30**	0.0.693	0.960
Mean Time to Plan	Level of Control	1	11	20.70**	0.653	0.985
	N WASMs	1	10	19.26**	0.794	0.998
	Level of Control by N WASMs	1	10	0.71	0.125	0.139
Mean Time to Complete	Level of Control	1	11	490.81**`	0.978	1.000
	N WASMs	1	10	6.89*	0.579	0.817
	Level of Control by N WASMs	1	10	3.56	0.416	0.524

N = 12

SD Time on Target Error (i.e., how close the attacks were to each other) was significantly affected by level of control, number of WASMs/targets, and their interaction. An examination of the means in Table 3 show that time between attacks was greater for the manual versus cooperative control condition and generally increased as the number of WASMs/targets increased.

Significant effects were observed for both *Time to Plan* and *Time to Complete* for level of control and number of WASMs/targets. *Time to Plan* was greater for manual control ($F(1, 11) = 20.70$, $p < .01$) and increased as the number of WASMs/targets increased ($F(2, 10) = 19.76$, $p < .01$). *Time to Complete* was *less* for manual control ($F(1, 11) = 490.81$, $p < .01$) and increased as the number of WASMs/targets increased ($F(2, 10) = 6.89$, $p < .01$). At first, it appears counterintuitive that *Time to Complete* was lower for the manual versus the cooperative control mode. However, it should be noted that in the manual control mode, target authorization and attack occur separately for each WASM/target combination and once authorization has occurred, the WASM takes a direct flight path to the target. In the cooperative control mode the attack does not occur till all target/WASM combinations have been authorized and it is necessary for some WASMs to employ longer flight paths to enable simultaneous attack.

3.2. Confidence Ratings

Means and standard deviations for participant's confidence in their target identification decisions are summarized in Table 5 and results of the repeated measures analysis of variance are summarized in Table 6. Although there was a trend toward greater confidence for decisions made using the cooperative control mode, this trend was not significant. It should be noted that the observed power for this test was low, suggesting that if a larger sample were tested the effect might reach statistical significance. Mean confidence level was related significantly to the number of WASMs/targets. An examination of the means showed a general trend toward lower confidence as the number of WASMs increased, especially for the manual control mode. Although confidence ratings varied, they were in the "fairly confident" to "very confident" range for all level of control by number of WASMs/targets combinations, even for the manual control mode with 16 WASMs/targets.

Table 5. Means and Standard Deviations: Confidence Ratings

		Cooperative Control		Manual	
Score	N WASMs	Mean	SD	Mean	SD
Confidence	4	4.97	0.096	4.97	0.096
	8	4.97	0.096	4.63	0.702
	16	4.63	0.481	4.33	0.550

N = 12

Table 6. Repeated Measures Analysis of Variance Multivariate Test Results: Mean Confidence Ratings

Score	Effect	df1	df2	F	Partial Eta^2	Observed Power
Mean Confidence	Level of Control	1	11	3.25	0.228	0.378
	N WASMs	2	10	9.52**	0.656	0.924
	Level of Control by N WASMs	2	10	1.61	0.244	0.263

**p≤ .01

3.3. Workload

3.3.1. Subjective Workload

Subjective workload was measured using the NASA TLX. As previously discussed, the NASA TLX has 6 subscales that are combined to create an overall workload index. Table 7 summarizes the means and standard deviations for the subscales and total score. Table 8 summarizes the results of the repeated measures analyses of variance. Examination of the means in Table 7 revealed a consistent trend toward increased workload going from the cooperative control mode to manual control mode and from 4 to 8 to 16 WASMs. As summarized in Table 8, this trend was statistically significant for the Total workload score and for all of the NASA TLX scales except Physical workload.

Table 7. Means and Standard Deviations: Subjective Workload Scores

		Cooperative Control		Manual	
Score	N WASMs	Mean	SD	Mean	SD
Mental Workload	4	11.44	9.98	25.69	15.46
	8	16.66	10.58	38.47	18.38
	16	22.50	14.57	50.69	21.46
Physical Workload	4	8.61	4.31	12.91	11.35
	8	9.02	5.14	15.27	14.31
	16	11.25	8.79	18.61	18.42
Temporal Workload	4	13.47	10.45	29.44	19.85
	8	15.27	10.19	40.27	20.14
	16	21.38	13.04	56.38	21.78
Performance Workload	4	10.69	8.91	27.77	16.89
	8	11.38	10.91	36.80	16.27
	16	13.88	10.90	44.02	20.66

Effort	4	13.61	10.09	27.91	18.43
Workload	8	15.00	11.12	37.36	22.34
	16	20.97	16.39	48.88	25.36
Frustration	4	13.75	9.95	21.94	14.59
Workload	8	14.30	9.62	31.38	18.75
	16	17.91	8.67	42.77	21.86
Total	4	13.91	8.81	28.81	16.61
Workload	8	15.37	10.13	38.97	18.07
	16	21.20	12.63	51.15	20.25

N = 12

Table 8. Repeated Measures Analysis of Variance Multivariate Test Results: Subjective Workload Scores

Score	Effect	df1	df2	F	Partial Eta^2	Observed Power
Mental Workload	Level of Control	1	11	31.35**	0.740	0.999
	N WASMs	2	10	11.28**	0.693	0.960
	Level of Control by N WASMs	2	10	9.65**	0.659	0.928
Physical Workload	Level of Control	1	11	4.04	0.269	0.450
	N WASMs	2	10	2.02	0.288	0.322
	Level of Control by N WASMs	2	10	1.65	0.249	0.270
Temporal	Level of Control	1	11	33.16**	0.751	0.999

Workload	N WASMs	2	10	16.76**	0.770	0.995
	Level of Control by N WASMs	2	10	10.16**	0.670	0.940
Performance Workload	Level of Control	1	11	25.59**	0.699	0.996
	N WASMs	2	10	11.31**	0.693	0.960
	Level of Control by N WASMs	2	10	9.43**	0.654	0.922
Effort Workload	Level of Control	1	11	20.85**	0.655	0.985
	N WASMs	2	10	6.53*	0.567	0.796
	Level of Control by N WASMs	2	10	4.44*	0.470	0.622
Frustration Workload	Level of Control	1	11	11.21**	0.505	0.861
	N WASMs	2	10	7.70**	0.606	0.859
	Level of Control by N WASMs	2	10	10.12**	0.669	0.939
Total Workload	Level of Control	1	11	32.06**	0.745	0.999
	N WASMs	2	10	13.16**	0.725	0.980
	Level of Control by N WASMs	2	10	8.09**	0.618	0.877

**p≤ .01

3.3.2. Objective Workload via Physiologically-Based Cognitive Workload Assessment

Since the 6 task conditions were not repeated, an independent dataset, entirely separate from the training data, was not available to select the optimally-trained network

in the top-down feature reduction paradigm. Had data from repeated trials been available, the network that used both the fewest number of features and had the highest classification accuracy on the training data conditions (4 WASM/Cooperative and 16 WASM/Manual) in the repeated trials would have been reported as being the optimally-trained network (procedure used by Monnin & Estepp, 2009; Wilson, Estepp & Davis, in press). Instead, the optimally-trained network was selected as that which resulted in the highest percentage of the 4 WASM/Manual conditions being classified as low cognitive workload, resulted in a minimum accuracy of 95% on the validation set, and used the smallest number of features to achieve these two criteria. The use of maximally-classified low cognitive workload on the 4 WASM/Manual condition as a network selection metric will produce classification results that represent an upper bound on classification accuracy error (given that the 4 WASM/Manual condition was rated subjectively more cognitively demanding, as evidenced by the NASA TLX total workload scores, than all of the cooperatively-controlled conditions, but produced the lowest NASA TLX workload score of the manually-controlled conditions) while still maintaining high generalization in classifying the validation set data (using a minimum 95% correct classification accuracy).

The results presented in Figure 2 show the percentage of each data class that was classified as being from a low-workload task demand condition, based on the selection of results from the optimally-trained network (as described above). Of the 12 participants for which performance and subjective workload data were presented, classification accuracy data from only 10 of the 12 participants is reported due to low quality of the physiological data (missing data channels, high impedance, poor signal-to-noise ratio, etc.). Error bars on each of the reported data points represent standard error of the mean.

In order to assess how well the physiological workload analysis results in Figure 2 compared with the task performance metrics and subjective workload, correlation coefficients were calculated between the physiological workload analysis results (% of each data class calculated as low workload) and both the task performance and subjective workload metrics (NASA TLX). Table 9 summarizes the results of the correlation analysis. Four of the objective task performance measures involving efficiency (Mean Time on Target Error, SD Time on Target, Time to Plan, and Time to Complete) were

strongly correlated (±0.75 of greater) with the physiological measures. Negative

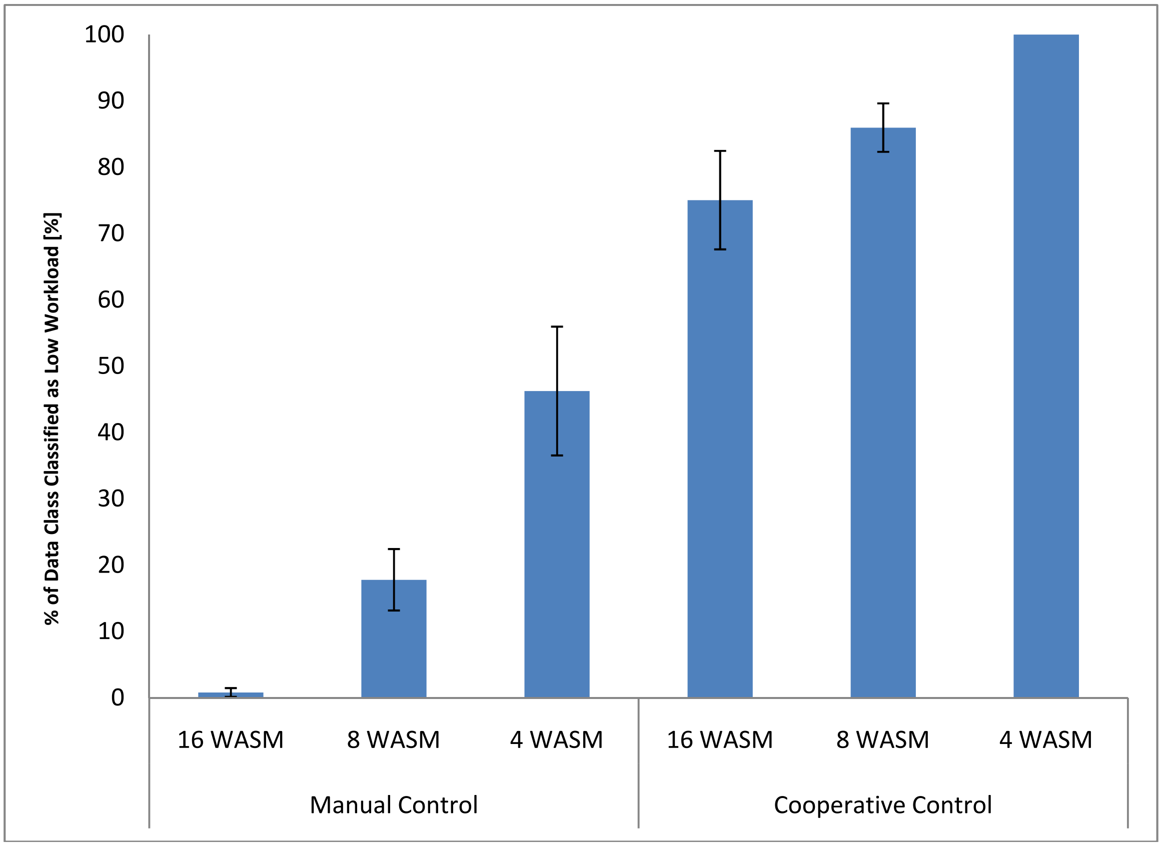

Figure 2. Results of ANN cognitive workload analysis, presented as percentage of each data class that was classified as being from a low workload condition. Error bars are standard error of the mean across participants (N = 10). Note that the standard error for the 4 WASM. Cooperative condition is 0, given that the ANN classified the data from the 4 WASM, Cooperative condition as 100% low workload for all 10 participants.

correlations between physiological workload and Mean Time on Target Error, SD Time on Target, Time to Plan indicated that as the proportion of low ratings of workload (proportion indicating the task was easy) increased, efficiency also increased (i.e., time required and variability in time required to complete a task decreased). The positive correlation between physiological workload and Time to Complete indicated that low ratings of workload were associated with taking longer to complete the task,

More striking, however, were the very strong correlations between physiological workload and the subjective NASA TLX workload scores. The correlations between objective and subjective workload scores ranged from -.96 to -.99 with a mean of -.98.

Table 9. Correlation between Results of Cognitive Workload Analysis and Task Performance Measures.

	Cooperative Control			Manual Control			
Measure	4 WASMs	8 WASMs	16 WASMs	4 WASMs	8 WASMs	16 WASMs	Correlation
% Classified as Low Workload	100.00	85.97	75.04	46.25	17.79	0.80	---
Mean Time on Target	494.01	488.58	540.15	573.85	446.71	552.56	-0.13
Mean Time on Target Error	2.04	1.31	6.59	---	---	---	-0.75
SD of Time on Target	2.25	1.46	9.09	10.17	17.58	27.44	-0.95
Time to Plan	22.47	36.01	70.16	39.40	61.26	105.25	-0.78
Time to Complete	117.22	124.63	148.10	63.06	65.65	74.96	0.77
# of High Priority Targets Attacked	3.33	6.67	12.31	3.28	6.56	12.53	-0.39
# of Low Priority Targets Attacked	0.67	1.33	3.69	0.69	1.33	3.42	-0.33
Mental Demand	14.44	16.67	22.50	25.69	38.47	50.69	-0.97
Physical Demand	8.61	9.03	11.25	12.92	15.28	18.61	-0.98
Temporal Demand	13.47	15.28	21.39	29.44	40.28	56.39	-0.98
Performance	10.69	11.39	13.89	27.78	36.81	44.03	-0.99
Effort	13.61	15.00	20.97	27.92	37.36	48.89	-0.99
Frustration	13.75	14.31	17.92	21.94	31.39	42.78	-0.96
Total Workload	13.92	15.37	21.20	28.81	38.97	51.16	-0.99
Confidence	4.97	4.63	4.33	4.97	4.97	4.63	-0.11

3.4. Post-Test Questionnaire

Following completion of the test trials, participants completed a post-study questionnaire regarding their experience. They rated ease with which they were able to use the operator interface to identify targets and their ability to classify the priority level of targets using the WASM interface. Both ratings were on a 5 point scale: 1 – poor, 2 – fair, 3 – good, 4 – very good, and 5 – excellent. Although ratings for ease of use and ability to classify the target priority level varied, the mean ratings for both approached "very good." Rating for ease of use ranged from 3 to 5 with a mean of 3.92; those for ability to classify the target priority level ranged from 2 to 5 with a mean of 3.83.

Participants then rated their ability to perform a simultaneous attack using the

cooperative control and manual control modes for the 4 and 16 WASM/target conditions. Ratings were on a five point scale: 1 – poor, 2 – fair, 3 – good, 4 – very good, and 5 – excellent. Means and standard deviations are summarized in Table 10 and results of the analysis of variance in Table 11. Inspection of the means showed a strong trend toward lower ratings of ability to perform a simultaneous attack for the manual control mode and for the 16 WASM/target condition. The effect was especially pronounced for the manual mode condition with 16 WASMs/targets (mean = 1.5).

Table 10. Means and Standard Deviations: Ability to Perform Simultaneous Attack

		Cooperative Control		Manual	
Score	N WASMs	Mean	SD	Mean	SD
Ability to Perform Simultaneous Attack	4	4.83	0.389	4.17	0.178
	16	3.83	0.835	1.50	0.674

N = 12

Table 11. Repeated Measures Analysis of Variance Multivariate Test Results: Subjective Evaluation of Ability to Perform Simultaneous Attack

Score	Effect	df1	df2	F	Partial Eta^2	Observed Power
Ability to Perform Simultaneous Attack	Level of Control	1	11	66.00**	0.857	1.000
	N WASMs	1	10	61.90**	0.849	1.000
	Level of Control by N WASMs	1	11	28.94**	0.725	0.998

N = 12

Participants had the opportunity to provide open-ended comments regarding the WASM interface and procedures. Seven of the 12 participants made one or more comments. These focused on ways to improve the manual mode and the interface design. Suggestions regarding the manual mode included adding the ability to insert waypoints and timing points to improve simultaneous attack. Suggestions regarding the interface design focused on providing multiple data input options (e.g., keyboard, voice) in addition to the mouse and using a larger screen or multiple screens.

4.0 DISCUSSION

Participants were able to acquire and attack nearly all of the targets even under the most demanding condition, that is, manual control of 16 WASMs. As expected, unaided operators were not able to achieve simultaneous attack of the targets as efficiently as the cooperative controller. Time between attacks was greater for the manual versus cooperative control mode and generally increased as the number of WASMs/targets increased. The decrement in performance efficiency between the manual and cooperative control modes is important under the circumstance when it is crucial to limit the amount of time an adversary has to respond to a first attack. Even in the least demanding condition involving 4 WASMs/targets, participants' ability to manually perform a near simultaneous attack was degraded compared to the cooperative control mode. These results also are reflected in participants' objective (physiological) and subjective (NASA TLX) workload and their ability to perform a near simultaneous attack.

The physiologically-based measure of workload was strongly related to both task performance efficiency and self-assessment of workload. Four of the objective task efficiency measures were related to the objective measure of workload and indicated that lower workload was associated with greater task efficiency. The very strong correlations between physiological workload and the subjective workload ratings (mean $r = -.98$) affirmed the construct validity of the subjective measure. Workload assessment as employed in this study was merely a post-hoc diagnostic tool that examined the effects of task complexity on performance. Given the resource requirements needed to measure physiological workload (i.e., equipment, administration, expertise, data analysis), the

simpler subjective questionnaire measure was an acceptable substitute in the current study. However, real-time workload assessment is needed where the objective is to provide the operator adaptive-aiding when warranted by task demands. Subjective post-hoc measures such as the NASA-TLX are inadequate for this objective. Wilson and Russell (2003b, 2007) demonstrated the utility of psychophysiological measures for monitoring cognitive workload and determining if and when system intervention should be provided to assist the operator and improve system performance. Wilson and Russell used real-time psychophysiological workload assessment and adaptive aiding to modify the UAV operator task to reduce cognitive demands when the operator was in a state of high cognitive workload. Adaptive aiding was accomplished by reducing the velocity of the UAVs and modifying the way vehicle health status messages were displayed. Additional studies should be conducted with higher expected workload levels to determine whether this relationship between the two workload assessment methods is observable throughout a wider range of task difficulty. Studies also are needed to further examine the utility of psychophysiologically-determined adaptive aiding for a broader range of operator tasks (e.g., route planning, task prioritization, target identification). In addition to workload, it is suggested that related studies employ additional measures to determine the cooperative controller (or associated decision aids) effect on operator situation awareness.

Additional studies are needed to examine other factors that may affect performance differences between the manual and cooperative control modes. For example, the extent to which targets are clustered (or dispersed) in the search area may affect the relative efficiency of the manual and cooperative control modes. Also, it would be informative to examine additional numbers of WASMs/targets (1, 2, 3, … n) across different priorities and/or attack sequences to better determine performance differences between the manual and cooperative control modes, as well as different strategies for using or combining manual and cooperative control modes.

These studies could examine the impact of adjustable automation on task management preferences and performance. Adjustable automation would allow the operator to change the automation level in response to changes in task requirements and workload. The operator’s decision to use automation can be affected by task complexity

which could be manipulated by varying the task demands (e.g., number assets they have to manage, number of targets), availability of decision aids (e.g., automatic router, target prioritization), the addition of other tasks (e.g., interacting with other manned and unmanned assets), time constraints, and mission goals (e.g., synchronization of attacks), and the operator’s level of experience/ability. The tendencies discovered could prove useful in developing effective adaptive interface techniques for future human-machine systems.

REFERENCES

Barbato, G. (2000). Uninhabited combat air vehicle controls and displays for suppression of enemy air defenses. *CSERIAC Gateway. 11 (1),* 1-4.

Clough, B. T. (2002). *UAV swarming? So what are those swarms, what are the implications, and how do we handle them?*, AFRL-VA-WP-TP-2002-308. Wright-Patterson AFB, OH: Air Force Research Laboratory, Air Vehicles Directorate, Control Sciences Division.

Goraydin, O. (2003). *Analysis of cooperative behavior for autonomous wide area search munitions.* Unpublished Master's thesis. Wright-Patterson AFB, IH: Air Force Institute of Technology.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA TLX (task load index): Results of empirical and theoretical research in human mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). New York: Elsevier.

Jasper, H. H. (1958). The 10–20 electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Krizo, E. J., Wilson, G. F., & Russell, C. A. (2005). NuWAM: The development of a modular, real-time augmented cognition system. *Proceedings of the 11th International Conference on Human-Computer Interaction: Vol. 11. Foundations of augmented cognition* [CD-ROM]. Mahwah, NJ: Erlbaum.

Wilson, G. F., Estepp, J. R. & Davis, I. E. (2009). A comparison of performance and psychophysiological classification of complex task performance. *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society,* San Antonio, TX.

Prieditis, A., Dalal, M., Arcilla, A., Groel, B., Van Der Bock, M, & Kong, R. (2004). *SmartSwarms: UAVs that think.* Lockheed Divisions, Inc., Davis, CA.

Ruck, D. W., Rogers, S. K., & Kabrisky, M. (1990). Feature Selection using a Multiplayer Perceptron, *Journal of Neural Network Computing, 2,* 40-48.

Scerri, P., Liao, E., Lai, J., Sycara, K., Xu, Y., & Lewis, M. (2005). Coordination very large groups of wide area search munitions. In D. Grundel, R. Murphey, & P. M. Partdalos (Eds.), *Series on Computers and Operations Research, vol. 4, Theory*

and Algorithms for Cooperative Systems (pp. 1-31). World Science Publishing Company.

Schumacher, C., Chandler, P. R., Rasmussen, S. J. (2002). *Task allocation for wide area search munitions,* AFRL-VA-WP-TR-2002-301. Wright-Patterson AFB, OH: Air Force Research Laboratory, Air Vehicles Directorate, Control Sciences Division.

Schumacher, C., Chandler, P. R., Rasmussen, S. J., & Walker, D. (2003). *Task allocation for wide area search munitions with variable path length,* AFRL-VA-WP-TR-2003-306. Wright-Patterson AFB, OH: Air Force Research Laboratory, Air Vehicles Directorate, Control Sciences Division.

Wilson, G. F., Estepp, J. R. & Davis, I. E. (2009). A comparison of performance and psychophysiological classification of complex task performance. *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society,* San Antonio, TX.

Wilson, G. F., & Russell, C. A. (2003a). Operator functional state classification using multiple psychophysiological features in an air traffic control task, *Human Factors*, *45,* 381–389.

Wilson, G. F., & Russell, C. A. (2003b). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors, 45,* 635–643.

Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an unmanned air vehicle task using psychophysiologically determined adaptive aiding. *Human Factors, 49,* 1005-1018.

APPENDIX

Study Questionnaires

Demographic Data Questionnaire

Participant ID: ___________

1. Age: ___________

2. Gender (circle one) Male Female

3. Describe your general health (circle one):

   Poor Fair Good Very Good Excellent

4. How would you assess your overall feeling of wellbeing this morning/afternoon (circle one)?

   Poor Fair Good Very Good Excellent

5. Do you have any practical experience working in a simulation type environment?

   If yes explain:

6. Do you play any type of computer/video games? Yes No
   a. If you answered "Yes," what types do you play? (circle all that apply)

   Action/Adventure _____ Role Playing _____
   Other (specify) _____________

   b. Do the computer/video games you play require you to do visual search tasks (i.e., locate/identify objects or targets)? Yes No

7. Is your visual acuity correctable to 20/20? Yes No

8. Do you have any problems with your peripheral vision? Yes No

9. Are you color blind? Yes No

10. Are you aware you may withdraw from this study at any time? Yes No

11. Are you aware that your participation is strictly confidential? Yes No

Wide-Area Search Munitions
Confidence Questionnaire

Participant:__________________________
Date: ______________________________

Four WASM Condition

1.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

2.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

3.) Run: _____________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

4.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

Eight WASM Condition

Participant:__________________________
Date: ______________________________

1.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

2.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

3.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

4.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

Sixteen WASM Condition

Participant:__________________________
Date: ______________________________

1.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

2.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

3.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

4.) Run: ______________________________

How confident are you that you attacked the appropriate high and low priority targets? Please rank your decision between 1 and 5 with (1- not at all confident, 2- slightly confident, 3- moderately confident, 4 - fairly confident, 5 – very confident).

NASA-TLX Instructions and Questionnaire

We are not only interested in assessing your performance but also the experiences you have during the experimental trials. Right now we are going to describe the technique that will be used to examine these experiences. In the most general sense we are examining the “Workload" you experience. Workload is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of workload may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you feel. Physical components of workload are relatively easy to conceptualize and evaluate. However, mental components of workload may be more difficult to measure.

Since workload is something that is experienced individually by each person, there are no effective "rulers" that can be used to estimate the workload of different activities. One way to find out about workload is to ask people to describe the feelings they experienced. Because workload may be caused by different factors, we would like you to evaluate several of them individually rather than lumping them into a single, global evaluation of overall workload. This set of six rating scales was developed for you to use in evaluating your experiences during different tasks. (Hand scale sheet on top of explanations to participant)

Please read the descriptions of the scales carefully. If you have a question about any of the scales in the table, please ask me about it. It is important that they be clear to you. You may keep the descriptions with you for reference during the experiment.

(Stop here, read detailed subscale explanations while participant reviews the scale sheet/explanations)

After performing each task, you will evaluate it by marking each scale at the point that matches your experience. Each line has two endpoint descriptors that describe the scale. Note that "performance" goes from “good" on the left to “poor" on the right. This order has been confusing for some people. Mark the desired location. Please consider your responses carefully in distinguishing among the task conditions. When rating each task, only reflect on the one you have just completed. Consider each trial in isolation, that is, do not compare it to prior experiences. Also, please consider each scale individually. Although the definitions may be similar for two or more scales, try to distinguish them from each other based on my explanations and the definitions that you may refer to throughout the experiment- even when rating them.

Your ratings will play an important role in the evaluation being conducted, thus, your active participation is essential to the success of this experiment, and is greatly appreciated!

NASA Task Load Index (TLX) Paper and Pencil Package Version 1.0 (1986). Moffett Field, CA: Human Performance Research Group, NASA Ames Research Center.

NASA –TLX

NASA TLX RATING SCALE DEFINITIONS		
Title	**Endpoints**	**Descriptions**
MENTAL DEMAND	Low High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
EFFORT	Low High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
PERFORMANCE	Good Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
FRUSTRATION LEVEL	Low High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

NASA –TLX

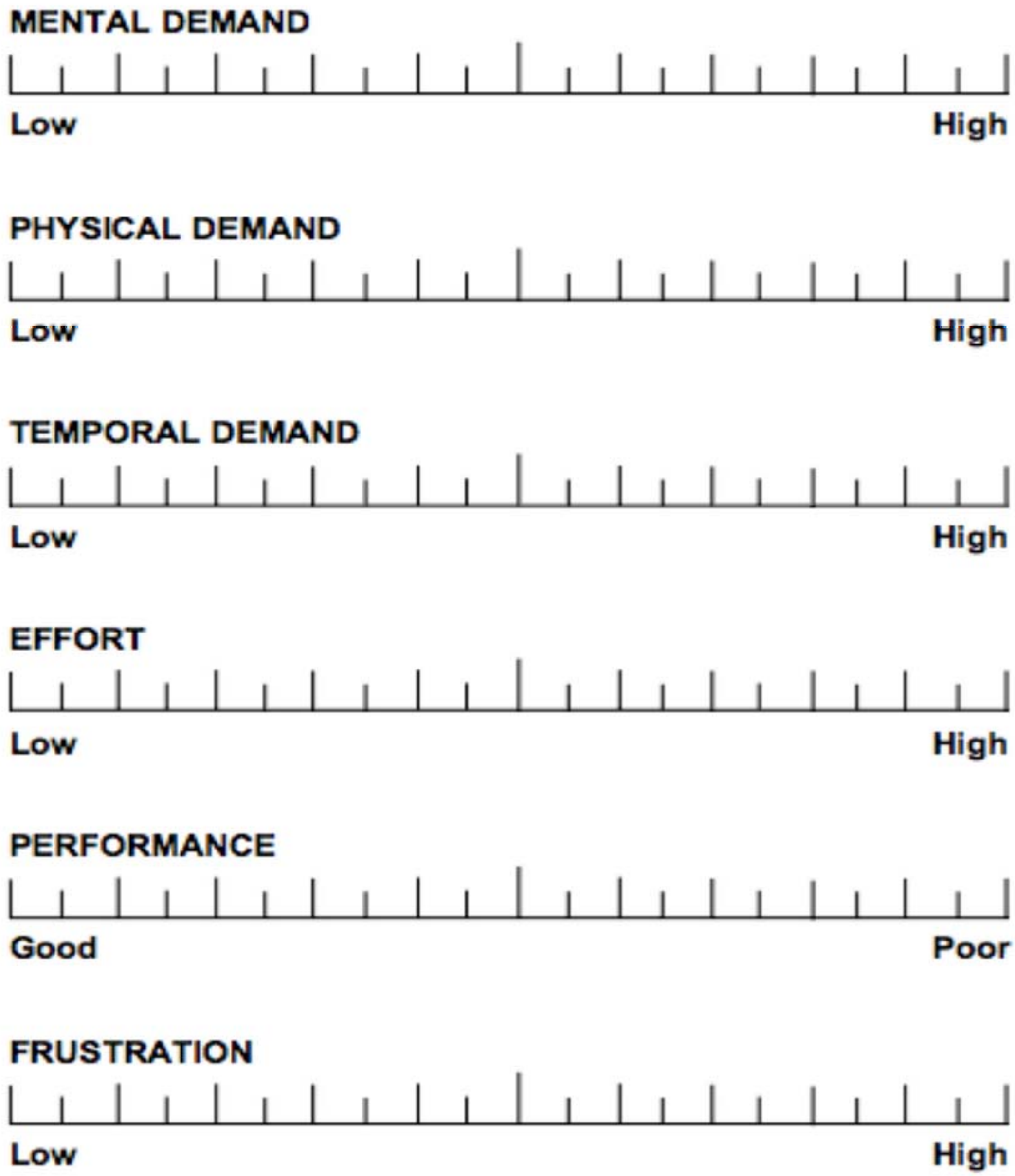

Figure A-1. **NASA TLX rating scales**.

Post-Test Interview Questions

Subject: ________

Date: _______

1. How would you rate the ease with which you were able to use the operator interface to identify the targets? (circle one)

   Poor Fair Good Very Good Excellent

   If your answer to #1 was "poor" or "fair," what factors affected your rating?

2. How did the operator interface affect your ability to classify the target's priority level? (circle one)

   Poor Fair Good Very Good Excellent

3. When using the manual mode of attack in the 4 target condition, rate your perceived ability to accomplish a near simultaneous attack with an equal number of WASMs, (circle one)

   Poor Fair Good Very Good Excellent

4. When using the cooperative control mode of attack in the 4 target condition, rate your perceived ability to accomplish a near simultaneous attack with an equal number of WASMs, (circle one)

   Poor Fair Good Very Good Excellent

5. When using the manual mode of attack in the 16 target condition, rate your perceived ability to accomplish a near simultaneous attack with an equal number of WASMs, (circle one)

   Poor Fair Good Very Good Excellent

6. When using the cooperative control mode of attack in the 16 target condition, rate your perceived ability to accomplish a near simultaneous attack with an equal number of WASMs, (circle one)

   Poor   Fair   Good   Very Good   Excellent

7. Were you able to identify all predefined targets of interest in the video? (circle one)   Yes   No

   If no, explain:

8. Please provide any additional comments below: